**Университет ИТМО**

Кафедра информатики и прикладной математики

**Лабораторная работа №2**

Выполнила:

Бриль Марина P4217

# 

# 

### **Структура исходных данных**

The instances were drawn randomly from a database of 7 outdoor images. The images were handsegmented to create a classification for every pixel.

Each instance is a 3x3 region.

Attribute Information. Float values:

1. region-centroid-col: the column of the center pixel of the region.

2. region-centroid-row: the row of the center pixel of the region.

3. region-pixel-count: the number of pixels in a region = 9.

4. short-line-density-5: the results of a line extractoin algorithm that

counts how many lines of length 5 (any orientation) with

low contrast, less than or equal to 5, go through the region.

5. short-line-density-2: same as short-line-density-5 but counts lines

of high contrast, greater than 5.

6. vedge-mean: measure the contrast of horizontally

adjacent pixels in the region. There are 6, the mean and

standard deviation are given. This attribute is used as

a vertical edge detector.

7. vegde-sd: (see 6)

8. hedge-mean: measures the contrast of vertically adjacent

pixels. Used for horizontal line detection.

9. hedge-sd: (see 8).

10. intensity-mean: the average over the region of (R + G + B)/3

11. rawred-mean: the average over the region of the R value.

12. rawblue-mean: the average over the region of the B value.

13. rawgreen-mean: the average over the region of the G value.

14. exred-mean: measure the excess red: (2R - (G + B))

15. exblue-mean: measure the excess blue: (2B - (G + R))

16. exgreen-mean: measure the excess green: (2G - (R + B))

17. value-mean: 3-d nonlinear transformation

of RGB. (Algorithm can be found in Foley and VanDam, Fundamentals

of Interactive Computer Graphics)

18. saturatoin-mean: (see 17)

19. hue-mean: (see 17)

Classes:

* brickface,
* sky,
* foliage,
* cement,
* window,
* path,
* grass.

### **Результаты**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **experiment number** | **train set size** | **decision tree accuracy** | **random forest accuracy** |
| **0** | 1 | 60.0 | 88.095238 | 92.500000 |
| **1** | 1 | 70.0 | 89.523810 | 90.634921 |
| **2** | 1 | 80.0 | 87.619048 | 91.904762 |
| **3** | 1 | 90.0 | 86.666667 | 91.428571 |

### **Вывод**

В ходе лабораторной работы были получены практические навыки работы с методом деревьев решений на практических примерах с использованием языка программирования python и библиотеки sklearn.

Были использованы классификаторы Dicision Tree Classifier и Randrom Forest Classifier, показавшие хорошую точность предсказаний (88 - 100%) с различными по размеру обучающими выборками. Тем не менее Randrom Forest Classifier показал лучший результат по сравнению с Dicision Tree Classifier в каждом эксперименте